NASA TECHNICAL NOTE

NASA TN D-4572

c.1

NASA TN D-4572

CODING SCHEMES FOR RUN-LENGTH INFORMATION BASED ON POISSON DISTRIBUTION

by W. W. Happ

Electronics Research Center
Cambridge, Mass.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JUNE 1968

NASA TN D-4572

CODING SCHEMES FOR RUN-LENGTH INFORMATION BASED ON POISSON DISTRIBUTION

By W. W. Happ

Electronics Research Center
Cambridge, Mass.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

CODING SCHEMES FOR RUN-LENGTH INFORMATION BASED ON POISSION DISTRIBUTION

By W. W. Happ
NASA Electronics Research Center

SUMMARY

The Shannon-Fano-Huffman Redundancy Reduction Procedure is extended to one-parameter and two-parameter Poisson distribution resulting from run-length information streams. Figures-of-merit are established to compare Huffman coding to fixed-word-length binary coding on the basis of (1) average number of coding digits per message, (2) signal-to-noise ratio in analog-to-digital conversion due to bit errors, and (3) design criteria primarily of interest to the system designer and not readily amenable to quantitative analysis. To ascertain whether or not Huffman coding leads to a unique and optimum code, other equally good optimum solutions were obtained. Optimization and uniqueness criteria are examined and applications to data compression are discussed.

INTRODUCTION

The need for data compression or "compaction" is particularly acute when statistical information from deep space probes is transmitted. A series of important recent investigations (ref. 1) examined criteria for transmitting a small number of sample quantile, often referred to as percentage points, instead of all sample values. An alternative approach is proposed. The approach here presented has several disadvantages compared to the quantile system, specifically:

1. The code has variable length, with the inherent changes of errors in transmission wrecking the frame.

2. The code assumes "a priori" information on the statistical distribution, an assumption which is frequently not justified in practice.

3. The proposed approach has not been justified experimentally.

Despite these shortcomings, there are some significant similarities in the objectives, as well as in the statistically oriented features of these two investigations. For practical applications, the method described herein does not presently warrant comparison with that reported earlier, but should be regarded rather as a potential stepping stone for future investigations.

REVIEW OF PREVIOUS WORK

Huffman (refs. 2 and 3) defines a Minimum Redundancy Code as an ensemble or alphabet consisting of N members or letters with the following properties:

1. Each letter has a preassigned probability of occurrence $P(i)$ where $i = 1, 2, 3, \ldots N$ subject to normalization condition $S(i = 0, N)\ P(i) = 1$, where the summation $S(i = 0,\ N,)$ sums i from 1 to N.

2. Each letter is to consist of a number or length of coding digits $L(i)$, which are usually, but not necessarily, binary. The average message length $D = S(i = 0, N)\ P(i)\ L(i)$ is to be a minimum.

3. Subject to the normalization condition (1) and the minimization constraint (2) above, coding digits are to be assigned to generate N distinguishable letters, such that addition (or subtraction) of a digit from the end of a letter does not generate a new letter of the alphabet.

The Huffman procedure of constructing Minimum Redundancy Codes, is shown in part (a) of Table I for an alphabet with $N = 25$ and assigned values of $P(i)$. A figure-of-merit is $D = 4.24$, the sum of $P(i)\ L(i)$, that is, the average word length to assess the saving obtained by Huffman coding. This value must be compared with $D = 5$, for a $25 \leq 2^D$ letter conventional alphabet.

STATISTICAL PROPERTIES OF RUN-LENGTHS

A run of variable length r is defined if no event occurs during interval $1, 2, 3, \ldots r - 1$ followed by an event during interval r. Redundancy reduction uses information contained in the assumed or implied probability of occurrence of run of length r. The Poisson distribution describes a wide range of physical phenomena and Operations Research models. The simplest model is based on a single parameter, namely m $E(r)$, the mean occurrence or expected value of the run. It is assumed that r is always a non-negative integer, but m may or may not be an integer. The probability that exactly r events occur per run is denoted by

$$p(r, m) = m^r e^{-m} / r\,!$$

Assuming an expected value of run $m = 0.5$ and $m = 3$, representative values for $p(r, m)$ are listed in Table II. In evaluating experimental data, it is useful to obtain $p(r, m)$ experimentally and to calculate m using the following properties of the distribution:

TABLE I.
MINIMUM REDUNDANCY CODE

Input Data			(a) Minimum Code			(b) Quasi-Minimum Code			
i	N(i)	P(i) %	8 7 6 5 4 3 2 1	D	Code	L(i)	N(i)W(i)	P(i)L(i)	Code
E		13	N	.39	111				1111
	1		N			4	4	.52	
T		10	Z N	.30	110				1110
	1					4	4	.40	
A		7	N	.28	1011				1101
I			Z N Z	.28	1010				1100
O			N Z	.28	1001				1011
N			Z	.28	1000				1010
R			N	.28	0111				1001
S			Z N	.28	0110				1000
	6		N Z			4	24	1.68	
H		4	N Z	.16	0101				0111
	1					4	4	.16	
C		3	N Z	.15	01001				01101
D			Z	.15	01000				01100
F			N	.15	00111				01010
L			Z N	.15	00110				01010
M			N Z N Z	.15	00101				01001
U			Z	.15	00100				01000
	6					5	12	.90	
B		2	N	.10	00011				00111
G			N Z N Z	.12	000101				00110
P			Z Z	.12	000100				00101
W			N	.12	000011				00100
Y			Z N	.12	000010				00011
	5		Z			5	10	.50	
K		.5	N	.035	0000011				000101
J			Z N	.035	0000010				000100
Q			N Z	.04	00000011				000011
V			Z N	.04	00000010				000010
X			N Z	.04	00000001				000001
Z			Z	.04	00000000				000000
	6					6	6	.18	

TABLE II
PROBABILITY p(r, m) FOR RUNS OF LENGTH r AND
EXPECTED VALUE m

Binary	Huffman	p(r,1.3)	d (r, 0.5)	p(r, 3)	s(r, 3)
000	0	.60	.60	.05	.05
001	10	.30	.60	.15	.30
010	110	.075	.22	.22	.66
011	1110	.012	.05	.22	.88
100	11110	.002	.01	.16	.80
101	111110	.0002	.002	.09	.54
110	1111110	.00002	.0002	.06	.42
111	11111110	.000001	.00001	.03	.24

1. A maximum of p(r, m) occurs for $r = m^*$, where m^* is an integer equal to or next below m. If m is an integer, $p(m, m) = p(m - 1, m)$.

2. The standard deviation $S(r) = m^{1/2}$.

3. The mean deviation is $D(r) = 2m\, p(m^*, m)$.

Since $p(m^*, m)$ is frequently substantially greater than other values of p(r, m), it is often convenient to measure m by computing D(r) rather than S(r).

TRUNCATED RUN-LENGTH DISTRIBUTIONS

A Poisson distribution can be filtered or truncated at either end of the probability frequency spectrum, if runs below a given length or above a given length are eliminated. For example; the two-parameter probability density distribution

$$p(r, m, n) = 0 \text{ if } 0>r>n$$
$$= p(r, m) \,/\, G(n, m) \text{ if } n \leq r$$

is defined by supressing all runs of length less than n. The normalization function

$$G(n, m) = S(r = n, \infty)\, p(r, m) = S(m = 0, n)\, p(n,\ m-1),$$

that is

$$G(n, m') = \sum_{r=n}^{r=\infty} p(r, m') = \int_{o}^{n} p(n, m'-1)\, dm',$$

where $m' = E(r)$ is the expected value after truncation and $G(n, m)$ is the incomplete Gamma Function tabulated by Pearson (ref. 4), or is obtainable from tables of $p(r, m)$ computed by Molina (ref. 5).

Experimentally, the characteristic values of this two-parameter distribution are obtained from

$E(r) = m + n$, the expected value;
$S(r) = m^{1/2}$, the standard deviation; and
$D(r) = 2\, mp\, (m^*, m)$, the mean deviation.

The expected value is increased, since low values of r are omitted, while $S(r)$ and $D(r)$ are not affected by m and thus are well suited to numerical evaluation. The same approach is used for truncation of high values, or for band-pass and band-stop truncations.

HUFFMAN CODING OF POISSON DISTRIBUTED RUN-LENGTHS

An ensemble with 2^3 permissible run-lengths is assumed in Table II, and a possible Huffman code is assigned to each run. The probability density $p(r, m)$ is computed assuming a Poisson distribution with $m = 0.5$ and $m = 3$. The weighted run-length $d(r, m) = rp(r, m)$ is tabulated to compute the mean word length $D(m) = S(r = o, m)$ $d(r, m)$. Within the accuracy of the calculations $D(m) = m + 1$. This is a significant result in several respects:

1. The average information content is m for the message itself.

2. Shannon (ref. 6) showed that the average number of binary digits required for coding a message exceed the information content (binary measured) by a value which reflects the redundancy inherent in the coding scheme.

3. The average word length of Huffman-code Poisson-distributed run-length exceeds its information content by unity, independent of average message length.

4. The excess of average word length over average information content is not due to redundancy, but corresponds to unity information content per message, signaling the termination of a message.

5. The alphabet or message ensemble has an arbitrarily large number of messages, with each message being uniquely coded. Further, no coding sequence exists without a corresponding meaningful message. The ensemble is unique and non-redundant.

6. The combined information content, $m + 1$, consists of the message m augmented by a bit signaling the end of each message and equals the average word length. This accounts for the average word length providing conformation of the conclusion arrived at independently in (5) above that the code is unique and non-redundant.

In the example shown in Table II, a k-digit ($k = 3$) binary alphabet is compared with a run of length m as the expected value. Data compaction can be defined as the ratio of average word length binary $D(k)$ to average word length coded $D(m)$. For binary to Huffman coded run-length:

$$D(k) / D(m) = k / (m + 1)$$

In Table II for $m = 0.5$, then $D(k) / D(m) = 2$; and for $m = 3.0$, then $D(k) / D(m) = 0.75$

This shows that, for example, a five-fold compaction is feasible if $k = 8$ and $m = 0.6$.

ADEQUACY OF MATHEMATICAL MODEL

Confidence in the above results wanes as soon as the tenuous nature of the assumption is examined, upon which the mathematical model was formulated. Poisson distribution subjected to statistical analysis is based on several distinct models originating from a random arrival of a time sequence of events:

1. Given an average rate of occurrence, say m events per unit time, a uniform probability density m dt is assumed for event r, given the probability $p(r - 1, m)$. This yields the probability of obtaining a specified number of r events as a function of a variable time interval.

2. Given a fixed time interval in which an average of m events occurs, the stochastic variable r is the probable number of events likely to occur.

3. Both types of distribution can be modified by a non-zero, but constant, pulse duration during which no other event is acceptable to the counting mechanism.

For the run-length of primary interest in practical communication systems, a randomly arriving sequence of events is sampled at quantized time intervals. As the sampling period becomes small compared to mean occurrence of events, the above models become valid. Unfortunately, in the region of interest this condition is not satisfied. A sampled data output from a Poisson-distributed input can be obtained by z-transformation of the moment-generating function characterizing a Poisson distribution.

To the knowledge of the author, investigations on sampled data or on statistical properties of runs have not attacked this problem. A solution to the problem is of considerable practical interest and well within reach of presently available transform methods, from the vantage point of sampled data technique as well as from that of stochastic processes. It is questionable, however, if an investigation of this procedure will effectively serve to ascertain the adequacy of a mathematical model, unless the sensitivity of Huffman coding to deviations from ideal distributions is better understood.

Assuming that Huffman coding is highly sensitive to changes in the model, second-order approximations in formulating or modifying the model are warranted. Fortunately, it appears that even a rough approximation of the statistical model may be justified, since only minor changes in compaction ratio can be expected even by significantly deviating from optimum coding procedures.

In part (b) of Table I, a deviation from an optimum procedure is undertaken by grouping the letter of the alphabet. An additional restriction to Huffman coding is then imposed that each group defined by its probability must have the same run-length. A comparison from Table I is instructive:

	Minimum-Code	Non-Minimum
Average Word Length	D = 4.24	D = 4.36
Range of Runs	3, 4, 5, 6, 7, 8	4, 5, 6
Compaction Ratio	1.18	1.15

The insignificant difference in compaction ratio provides evidence that:

1. A quasi-minimum code does not deviate significantly from an optimum code.
2. Constraints were imposed on a simple model without changing significantly established figures-of-merit.
3. Other figures-of-merits, such as requirements on grouping or range of values, can provide useful information.

A similar approach can be taken to introduce desired changes in the Poisson distribution, for example, truncation, to explore variations in average word length.

TRADE-OFF AND OPTIMIZATION PROCEDURES

The feasibility of using models which contain additional constraints was explored by examining codes which differ from the Minimum Redundancy Code by an amount negligible in practice, but which exhibit useful properties not necessarily available from minimum codes. These codes will be termed "quasi-minimum" codes. Several questions arise regarding trade-off criteria and procedure, and properties of these codes enhancing their utilization.

1. Is it a "minimum" code? It is difficult to ascertain if the code derived by the Huffman technique is in fact a minimum. While no proof is as yet available, it appears plausible that Huffman's procedure, as well as the trade-off procedure to be developed in Table III, leads to a Minimum Redundancy Code.

2. Is there a unique code? Comparison of the codes in part (a) of Table I and part (3) of Table III shows that there are several codes which satisfy criteria defined by Huffman as an optimum code. To obtain one optimum code from another, it is necessary to employ successive application of the trade-off procedure given below.

3. What properties do minimum or quasi-minimum codes exhibit? Specifically, can features such as error detecting criteria or parity checks be incorporated into Minimum Redundancy Codes. Some statistical properties of these codes will be examined.

The quasi-minimum code developed in part (b) of Table I is summarized in part (a) of Table III. The number of letters in each group is denoted by N(i), and the corresponding probabilities listed under P(i) are the same as given in Table I.

A Minimum Redundancy Code or other optimum performance can then be obtained by:

1. Establishing a quasi-minimum code.

2. Establishing a trade-off procedure.

3. Repeating trade-off until an optimum is reached.

TABLE III
OPTIMIZATION PROCEDURE

Input Data		(a) Quasi-Minimum Code				(b) Minimum Code			
N(i)	P(i) %	L(i)	W(i)	N(i)W(i)	P(i)L(i)N(i)	L(i)	W(i)	N(i)W(i)	P(i)L(i)N(i)
1	13	4	4	4	.52	3	16	16	.39
1	10	4	4	4	.40	3	16	16	.30
6	7	4	4	24	1.68	4	8	48	1.68
1	4	4	4	4	.16	4	8	8	.16
6	3	5	2	12	.90	5	4	24	.90
5	2	5	2	10	.50	6	2	10	.60
6	.5	6	1	6	.18	7	1	6	.21
				64	4.36			128	4.24

To establish a quasi-minimum code, proceed as in part (a) of Table III:

1. Assume values for L(i), length of word for each category.

2. Assign a weight W(i) to L(i), such that two successive levels differ by a factor of 2.

3. Compute P(i) L(i) N(i) subject to the normalization constraint $N(i)\ W(i) \geq 2^{L(M)}$, where L(M) is the maximum value of L(i).

To establish a trade-off, proceed from part (a) of Table III to part (b) of Table III as follows:

1. Establish values for W(i) as before, and change L(i) by one unit subject to the normalization condition $N(i)\ W(i) \geq 2^{L(M)}$.

2. For every change in L(i), ascertain that the contribution to P(i) L(i) N(i) is negative before implementing it.

3. Repeat with (1) and (2) until an optimum is reached.

Rather than the derivation of any one optimum coding scheme, the significant conclusions to be drawn from Table III are:

1. The lack of sensitivity of average word length for near optimum codes;

2. The availability of step-wise optimization and tests for sensitivity for step-wise changes;

3. The existence of more than one code satisfying Huffman's criteria for uniqueness.

Quasi-optimum codes will be examined in the subsequent sections in view of their applicability to practical problems.

CODING FROM ASSUMED PROBABILITY SPECTRA

Any redundancy elimination scheme is based on utilization of information contained in the probability spectrum. Therefore, any figure-of-merit is only meaningful in terms of known or assumed data. To illustrate this, examine the alphabet based on the following probability spectra:

1. If all letters have the same probability, the Huffman coding procedure yields a binary code of uniform length, or letters at most differing by one digit.

2. If successive letters in an alphabet differ in probability by a constant ratio of $r = 2$, all letters differ by one digit in length and the mean length of the alphabet is $r^2(r-1)^{-2}$.

3. If all letters of an alphabet differ in probability by a constant ratio $r^k = 2$, any k letters will have the same number of digits.

The Poisson-distribution coded by the Huffman procedure yield several codes:

1. If the mean m does not exceed unity greatly, all letters differ by one digit; this arrangement will be termed Poisson-coding.

2. If the mean is $m >> 1$, then two regions should be distinguished: inside and outside the band $m \pm \sqrt{m}$.

3. In the region outside the band, Poisson coding gives an optimum scheme; in the inside region, Poisson coding is not desirable.

4. An optimum can be derived for the region outside the band by the optimization procedures discussed above. Alternately, an estimate of the average letter length can be obtained from ratios of successive probabilities.

5. For m = 1, the Poisson distribution approaches the normal distribution. By selecting a suitable coordinate system, the normal distribution has a mean of zero and a deviation of unity. Therefore, the sampling rate determines the deviation from the mean.

6. Assumptions made for the Poisson distribution do not necessarily apply to coding of normally distributed data. Specifically, caution is needed to ascertain the underlying assumptions to normally distributed data, before coding it by schemes developed for Poisson distributions.

Poisson-codes as defined above have distinct properties which can be exploited in several ways, namely, with respect to:

1. Subframe syncs;

2. Parity-type checks;

3. Error detecting schemes.

These will now be examined.

PROPERTIES OF POISSON-CODES

A series of interesting codes can be defined to describe runs with a known distribution spectrum. These codes differ from codes described by Huffman in several aspects, shown in Table IV:

1. The number of letters of the alphabet need not be finite.

2. Each code contains a zero at a fixed distance from one end, which is a "sync" to each subframe.

3. Each letter has a different length.

Various combinations of Binary and Poisson codes are possible, as shown in the example in Table IV under Poisson 2 and Poisson 4, each consisting of three parts:

1. A zero for sync;

2. A binary part, consisting of the binary number modulus 2^N;

TABLE IV
POISSON CODES

Decimal	Binary	Poisson # 1	Poisson # 2	Poisson # 4
0	0000	0	00	000
1	0001	10	01	001
2	0010	110	100	010
3	0011	1110	101	011
4	0100	11110	1100	1000
5	0101	111110	1101	1001
6	0110	1111110	1100	1010
7	0111	11111110	1101	1011
8	1000	111111110	11100	11000
9	1001	1111111110	11101	11001

3. A run of ones consisting of a Poisson code.

For the examples cited, the mean length for a Poisson distribution with $m = 1$ are:

Poisson # 1	L(P1) = 2.0
Poisson # 2	L(P2) = 2.9
Poisson # 4	L(P4) = 3.7
Binary	L(b) = 4.0

The distributions Poisson #2 and Poisson #4 differ from Poisson #1 in several respects. Specifically, they:

1. Do not conform to the criteria established by Huffman for an optimum code.

2. Correspond to no known distribution of data streams.

3. Do not satisfy the condition for quasi-minimum redundancy codes previously defined.

4. Contain ambiguity for "syncing" of subframe, since a sequence cannot be unequivocally recognized subsequent to a signal out of "sync".

It is not the purpose of these examples to derive optimum schemes, such as the coding scheme Poisson 1 in Table IV. Rather, the aim is to show that it is possible to define rational and valid criteria to compare coding schemes.

Further advantages of Poisson codes are:

1. The ratio of the number of zeros to the number of ones has an ascertainable expected value and an ascertainable standard deviation. Thus, two measurements are available for checks not too dissimilar in scope to parity checks.

2. A truncated Poisson-code gives in turn another Poisson code. Assume for example that no data involving 0, 1 and 2 are desired. Reducing all run-lengths by two and omitting 0, 1, 2 will again result in a Poisson distribution.

3. An error in "sync" or a bit error affects one or, at the most, two letters, and is comparable to errors in binary coding (ref. 7). Quantization errors are also comparable to those in binary coding.

The important aspect of the Poisson-code is its regularity and simplicity of structure. If applied to appropriate probability spectra, it provides not only significant compaction but also reliability features worth exploiting.

FIGURE-OF-MERIT: SIGNAL-TO-NOISE RATIO

In binary coding, the probability of a bit error (p) limits the signal-to-noise ratio for a long message:

$$R\ (K = \infty) = p^{-1}.$$

For a K-bit message

$$R(K) = p^{-1}\ (1-2^{-2K})^{-1}$$

Thus,

$$R(1) = (4/3)\ p^{-1}$$

The signal-to-noise ratio due to quantization (ref. 7) in binary coding is

$$Q(K) = 6(K+1)\ \text{db}.$$

In Huffman coding the signal-to-noise ratio R is similarly limited by

$$p^{-1} \leq R(K) \leq (4/3)p^{-1}$$

The quantization noise will be equal to the uncertainty in the run-length. For the Poisson distribution, the uncertainty is heavily weighted in favor of the more probable and shorter runs. The shortest runs are low accuracy measurements with an uncertainty of the order of 1/2 or 6 db for all measurements. The second moment of the Poisson distribution serves to calculate fluctuations in the run-length as $S(r = 0, \infty)\ p(r)\ (m-r)^2 = m$. The signal-to-noise level thus is m. Since the average run length $K = m + 1$, then:

$$Q(K) = 6\sqrt{K - 1} \quad \text{db}$$

For example, $m = 1/2$ yields $K = 1.5$ and $Q(K) = 4$ db. Figure-of-merit for signal-to-noise compared to quantization is:

$$Q(K, P) \ / \ Q(K, B) = (K - 1)^{1/2} \ / \ (K + 1) \approx K^{-1/2} .$$

Binary coding is far less sensitive to quantization errors, since all bits are weighted equally and only one bit is affected. In Poisson-coding, the shortest word fluctuates most and is most heavily weighted. Typically, assuming word lengths of 2.0 for both distributions, the signal-to-noise ratio is 30 percent better with binary coding.

FIGURES-OF-MERIT: AVERAGE WORD LENGTH

The average word length in a Poisson code is determined by the mean m of frequency distribution and is

$$L(P) = m + 1$$

The average word length for binary coding depends only on the desired accuracy and is

$$L(B) = -\ell n\ A.$$

For example, an accuracy of one part per thousand yields

$$A = 10^{-3} \text{ and } L(B) = 10.$$

For smaller values of m, the less deterministic are the data. In fact, if $m < 1$, then the standard deviation will exceed the mean. Poisson-coding is based upon, and exploits, the information contained in the statistical properties of data. This type of coding is therefore most effective when wide fluctuations in the data are anticipated, and data are needed with high accuracy. On the other hand, binary coding requires equal effort for highly probable as well as for rare events. Therefore, binary coding is advantageous for data streams of nearly equally probable events at uniform and low accuracy.

Consider the compaction ratio L(B) / L(P) for four cases:

	m = 1/2	m = 10
$A = 10^{-6}$	6.7	0.9
A = 5%	2.0	0.3

In summary, in binary coding, one pays for the highest accuracy with which the least probable data are given. In Poisson-coding, accurate data costs little, provided the need for accuracy occurs infrequently, and provided the coding is matched to the anticipated probability distribution.

APPLICATIONS TO MODULAR ARITHMETIC

A spectacular example of compaction resultant from variable-length codes can be found in coding the alphabet of modular arithmetic. An example is listed in Table V. Modular arithmetic has recently received attention in the Western technical literature, in view of emphasis given to the subject by computer designers in the Soviet literature.

In modular arithmetic, a number is written as a succession of remainders of modulus N(i), the jth prime number. For example,

(4, 2, 1) (modular) = 4(base 5) *2(base 3) *1(base 2)
= (29) (decimal notation).

Similarly

(6, 4, 2, 1) (modular) = 209 (decimal)

and

(10, 6, 4, 2, 0) (modular) = 2309 (decimal).

In Table V, the letter of the alphabet N(i) with associate probabilities P(i) is listed. From these input data, a quasi-minimum code is then obtained by computing L(i) and W(i) by the techniques developed for Table I.

The expected word length is calculated in Table VI assuming:

1. A uniform distribution of input data;

2. A Poisson distribution.

TABLE V
REDUNDANCY COMPARISON FOR MODULAR ARITHMETIC

N(i)	P(i)	L(i)	W(i)	W(i)N(i)	P(i)L(i)N(i)	K(i)	S(i)
2	50	3	16	32	3	.60	1.8
3	.33	4	8	24	4	.30	1.2
5	.20	5	4	20	5	.08	0.4
7	.14	6	2	14	6	.02	0.2
11	.09	7	1	11	7	.01	0.1
13	.07	8	.5	8	8		
17	0.6	9	.25	4	9		
19	0.5	10	.13	2	10		
21	0.5	11	.06	1	11		
23	0.4	12	.03	1	12		

TABLE VI
DATA COMPACTION FOR ALPHABET CODED FROM MODULAR ARITHMETIC

Modular Alphabet		Decimal Equivalent		Average Word Length		
Range	Digits	Range	Bits	Uniform	Poisson	Binary
4, 2, 1	3	29	2	12	4	5
6, 4, 2, 1	4	209	3	18	4	8
10, 6, 4, 2, 1	5	2309	4	25	4	12
	10	10^{10}	11	125	4	33

As before, the expected word length L(B) for binary coding is a function of the desired accuracy, namely L(B) = ℓnA. In Table VI, for A = 209 or .5% resolution, L(B) = 8. However, for Poisson coding the average word length L(P) is a function of the mean of Poisson distribution m.

Table VI compares typical cases. For example, A = 209 (decimal) = 6421 (modular) binary coding requires ℓn 209 or 8-digit coding. By the Huffman procedure given in Table V, the average path length is 3 + 4 + 5 + 6 = 18 using a uniform probability distribution. Using a Poisson distribution of mean m = 1 gives an average word length of L(P) = 4 as shown in Table V. The resultant compaction is L(B)/L(P) = 2.0.

A Poisson code for the modular alphabet combines features of both the high resolution of binary coding with the compaction inherent in Poisson coding. The compromise is based on the distribution of prime numbers, which is a compromise scale between linear presentation and logarithmic compression.

ALTERNATIVE APPROACHES

A detailed discussion of run-length coding is presented by Capon (ref. 8). Compaction techniques are examined by Cherry (ref. 9), Gouriet (ref. 10), as well as Becker and Lawton (ref. 11). A comprehensive analysis of Redundancy Reduction is presented by Gardenhire (ref. 12). A critical evaluation and comparison with these investigations is not attempted at this time.

REFERENCES

1. Anderson, T. O., Eisenberg, I., Lushbaugh, W. A. and Posner, E. C.: Demonstration of a Quantile System for Compression of Data from Deep Space Probe. IEEE Trans. AES-3, Jan. 1967, pp. 57-65.

2. Huffman, D. A.: A Method for the Construction of Minimum Redundancy Codes. Proc. IRE, Sept. 1952, pp. 1098-1101.

3. Jackson, W.: Communication Theory. Butterworth Scientific Publications, London, 1953.

4. Pearson, K.: Table of Incomplete Gamma Functions. H. M. Stationary Office, London, 1922.

5. Molina, E. C.: Poisson's Exponential Binomial Limit. Van Nostrand, New York, 1957.

6. Shannon, C. E.: A Mathematical Theory of Communications. Bell System Technical Journal, vol. 27, 1948.

7. Karp, S.: Noise in Digital-to-Analog Conversion Due to Bit Error. IEEE Trans. PG-Set 10, vol. 3, Sept. 1964, p. 124.

8. Capon, K.: A Probabilistic Model for Run Length Coding of Pictures. Transactions, Professional Group on Information Theory, IRE, December, 1959.

9. Cherry, C., et al: An Experimental Study of the Possible Bandwidth Compression of Visual Image Signals, Proc. IEEE, vol. 51, No. 11, Nov. 1963, pp. 1507-1517.

10. Gouriet, G. C.: Bandwidth Compression of a Television Signal: Proc. IEE, vol. 104, Part B, No. 15, May 1957, pp. 265-272.

11. Becker, H. D., and Lawton, J. G.: Theoretical Comparison of Binary Data Transmission Systems, Cornell Aeronautical Laboratory Inc., PAM. 621, 3841B3, May, 1958, Revised March, 1961.

12. Gardenhire, L. W.: Redundancy Reduction: The Key to Adaptive Telemetry. N. T. S. Conference, June, 1964.

National Aeronautics and Space Administration
Electronics Research Center
Cambridge, Massachusetts, January 1968
125-25-04-63

NASA-Langley, 1968 —— 7 C-32

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546

OFFICIAL BUSINESS

FIRST CLASS MAIL

POSTAGE AND FEES PAID
NATIONAL AERONAUTICS AND
SPACE ADMINISTRATION

02U 001 32 51 3DS 68150 00903
AIR FORCE WEAPONS LABORATORY/AFWL/
KIRTLAND AIR FORCE BASE, NEW MEXICO 87117

ATT MISS MADELINE F. CANOVA, CHIEF TECHNI
LIBRARY /WLIL/

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546